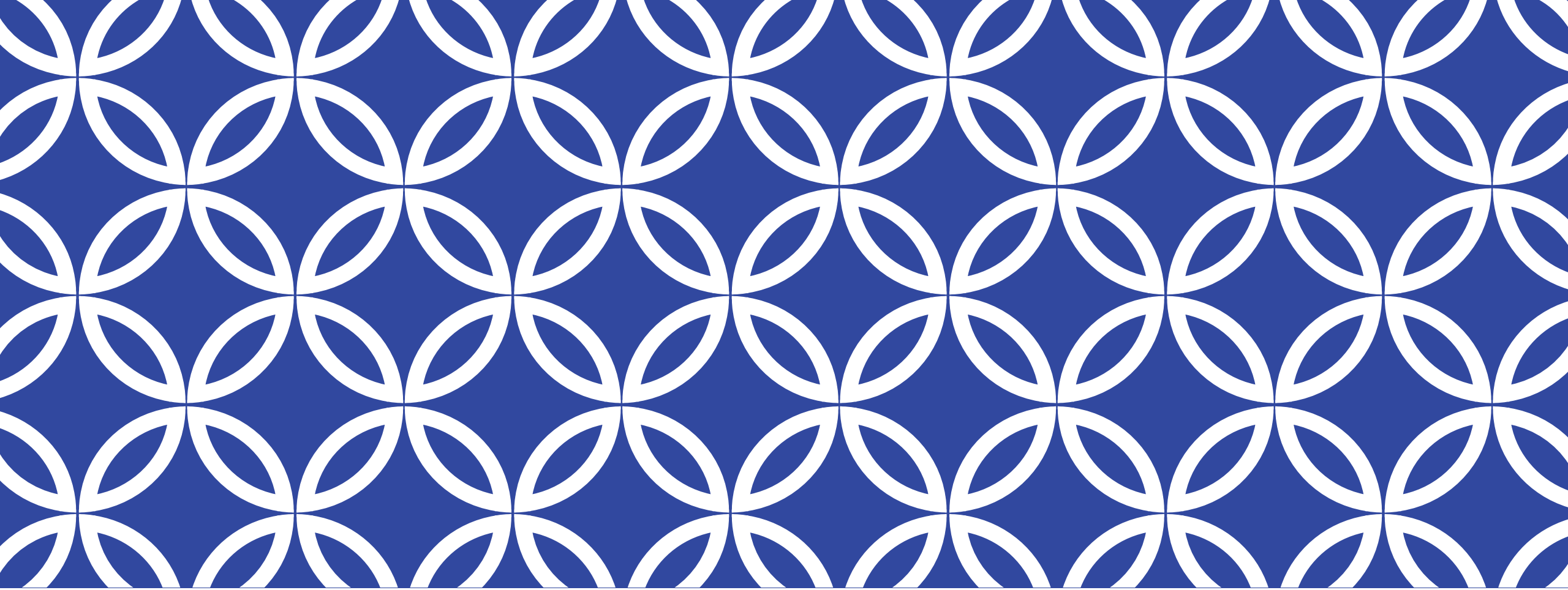

FINDING THE PERFECT CAR

A Springboard capstone project

PROBLEM

There are many places where one can search for used cars.

None can tell you about the safety ratings or reliability ratings of these cars.

None can tell you if a car has driven more or less than the average car of that brand, model and production year.

But this is very useful information to know when searching for a used car.

Having this information would make the client's web site unique, with added value for the users.

PROPOSAL

Combine three data sets to create a search function to find the perfect used car.

Use data from a web site for used cars (Ebay Kleinanzeigen), EURONCAP crash test data and reliability data (TÜV Auto Reports).

CHALLENGES

Crash test scores are saved in picture format

Solution – scrape the html-code and create a script that extracts the crash test rating from the picture names.

Car names are different

Solution – systemize the data so all three data sets describe the cars in the same way

Reliability data is registered biennially

Solution – create means from reports that overlap, so that every year has a distinct value

ABOUT THE WORK

Most of the time went to the data wrangling, as this was three not entirely clean data sets, where one was from a scraped web site. Fine-tuning the wrangling so that it can be used again and again should be prioritized if this project is adopted by the client's web site.

This project lays the foundation for further exploration into this subject. The result of the project is a working prototype of the new search. The new variables are:

Crash safety rating

Reliability data

Average mileage per brand, model and production year

EXAMPLE SEARCH

Search:

Only 5-star safety ratings

Fault rate below 10%

Price below 10 000 EUR

Only station wagons

A car with no unrepaired damage

Sorted with cars that have driven less than the average at the top

Top search result:

2012 BMW 520d Touring Edition with panorama sunroof, automatic transmission and heads up display

Price: 8 000 EUR

Driven 50 000 km (about half of the average for that car)

Reported fault rate for this model from 2012: 9,2%

CONCLUSION & RECOMMENDATION

The enhanced search allows the user to specify the parameters they value the most - safety and reliability rank among the highest prioritized properties in a car.

My client should consider working with automobile programs such as EURONCAP and TÜV auto reports to gather and use the ratings in their search application. There are several providers of reliability data, while EURONCAP is the assumed partner for crash test ratings.

Further exploration into other areas users care about (emissions? recyclability? future price prediction?) could be looked at next.